DOCUMENT RESUME

ED 152 420 PS 009 842

AUTHOR House, Ernest R.; Hutchins, Elizabeth J.
TITLE Issues Raised by the Follow Through Evaluation.
INSTITUTION ERIC Clearinghouse on Early Childhood Education, Urbana, Ill.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.d
PUB DATE 77
NOTE 19p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Demonstration Programs; *Early Childhood Education; Educational Accountability; Educational Assessment; Evaluation Methods; Federal Programs; Government Role; Primary Education; Program Effectiveness; *Program Evaluation; *Program Validation; *Research Design; *Research Problems; Statistical Bias
IDENTIFIERS *Project Follow Through

ABSTRACT

This paper presents a discussion of issues raised in the evaluation of Project Follow Through reported by Abt Associates. The paper suggests that many of the problems inherent in the design of both the program and the evaluation stem from the underlying assumption that one educational model could be found which would best alleviate the educational problems of the poor. The paper suggests that even when the original evaluation design was modified, substantial problems remained. The major issues and problems discussed in the paper include: (1) the belief in the existence of a best program; (2) the problem of relying on test scores; (3) the issue of program staff knowing the content of evaluation instruments and teaching to the test; (4) problems involved in designing or choosing valid instruments; (5) the existence of large intersite variations within the same models; (6) problems involved in implementing a particular model in varying sites; (7) statistical problems, particularly in the use of the analysis of covariance and the use of individual rather than class scores in the present evaluation; (8) problems involved in large scale experiments; (9) the fairness of the evaluation in terms of original intentions and later changes; (10) press coverage which tended to distort evaluation results, especially the invalid assumption that the basic skills programs were the most effective; and, (11) general questions of government policies which shaped the evaluation procedures and led to many of the subsequent problems. (BD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED152420

ISSUES RAISED BY THE FOLLOW THROUGH EVALUATION

Ernest R. House and Elizabeth J. Hutchins
University of Illinois

In April of 1977, Abt Associates, Inc. (AAI) released the long-awaited evaluation of the Follow Through program.* The AAI evaluation compared thirteen models of early childhood education, ranging from highly structured to open education approaches. The news media seized upon the findings as evidence that models labeled "basic skills" succeeded better than those labeled "cognitive" or "affective." The evaluation itself was strongly criticized by a panel of evaluation experts (House, Glass, McLean, and Walker, 1977). The evaluation is a porcupine of issues, some of which are discussed in this chapter.

Follow Through began in 1967 as a service program to continue the education of disadvantaged children, particularly those that had attended Head Start. It quickly ran into funding difficulties when the expected $120 million was reduced to only $15 million for the first year. Officials inside the federal bureaucracy decided to convert Follow Through into a "planned variation" experiment. That is, the government would support several types of early childhood models and eventually evaluate them

*Education at Experimentation: A Planned Variation Model, Volumes IV-A-D, Abt Associates Inc.: Cambridge, Mass., Richard B. Anderson, Project Director; Linda B. Stebbins, Deputy Project Director; Elizabeth C. Proper, Director of Analysis; April 15, 1977.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Ernest R. House
+Elizabeth J. Hutchins
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

PS009842

to see which worked best. This plan would enable Follow Through to continue.

Follow Through in its earliest planning stages was thought to be a program that could address change within institutions involving communities and families as well as schools. However, when Follow Through was designed as a planned variation experiment, the focus became less that of changing social institutions and more that of finding effective techniques of educating poor children in the existing school institution. Program planners chose to limit the program to finding techniques of schooling that would work better than traditional practices. In this way, the social service aspect of Follow Through was de-emphasized not only by the narrowness of the evaluation but also by the planners who chose the planned variation design.

The question to be answered by the evaluation was "what worked best" or "what worked most efficiently," as opposed to questions such as "how does it work" or "how can we make it work better." The history of the evaluation can be traced in excellent works by Haney (1977) and Elmore (1976). The policy which produced the evaluation has been analyzed by McLaughlin (1975) and House (1978).

The entire Follow Through program was begun in a social milieu of "can do." At that time in the sixties most educational reforms subscribed to the "big bang" theory of reform. They believed it was possible to discover a technique or a program that would "solve" a "problem" such as poor students failing to achieve in school. Not only could such a technique

be found but with some effort it could be disseminated all over the country, thus solving the social problem. Hence, the solution would be relatively cheap as well as affective.

Given such a belief, it became the mission of the federal government to discover techniques and disseminate them. First, the government had to find out what worked. Thus the strategy, clearly enunciated in the White House Conference on Education in 1967, (Elmore, 1976) was a matter of rectifying educators' ignorance. All would be well when the successful program was found.

The reformers ran into difficulties, however. Early evaluation results from Head Start and Title I, ESEA, indicated that the new reform programs were unsuccessful in raising the standardized test scores of the children involved. Federal officials interpreted this failure as inadequate variation and control over the programs. They concluded that efforts should be devoted to developing different programs and then systemmatically evaluating them. Hence, "planned variation," rather than natural variation, became a reform strategy. Follow Through was the first attempt at planned variation.

Sponsors, those developing the new models of early childhood education, and sites, school districts implementing the new models, were chosen by the Office of Education. Both sponsors and sites received funding from the federal government. At a meeting in Kansas City in 1968, sponsors and sites were matched to each other. Both development and implementation of the models, which were directed at poor children from kindergarten through third grade, began immediately.

From the viewpoint of the federal officials, particularly those within the Office of the Assistant Secretary for Planning and Evaluation inside the Department of Health, Education, and Welfare who had been instrumental in the planned variation conversion, evaluation was a critical part of the Follow Through plan. It would tell which model worked best and at what cost (Rivlin, 1971). Also, the federal planners had a particular idea of what evaluation should be--a massive, controlled experiment. It was a popular view of evaluation at that time, though not one universally shared within the evaluation community.

Consequently, the evaluation was set up as a large-scale experiment, with comparison groups assigned for each of the Follow Through classes. Comparisons would be made between the Follow Through and non-Follow Through classes. A huge, multi-million dollar contract was let to the Stanford Research Institute (SRI) to conduct the evaluation. SRI promised to evaluate all aspects of the program, including community involvement, institutional change, and so on.

However, when the evaluation began, SRI collected primarily standardized test scores. This upset many of the sponsors, and they protested vociferously. SRI assured them that the less tangible goals of their models would be assessed in addition to the more traditional outcomes as measured by standardized achievement tests. In fact, SRI began a serious effort to develop special measures appropriate to the expressed goals of the sponsor's models.

Meanwhile, the political pressure was intense in Washington to expand the number of sponsors and sites. Special interest groups like blacks and bilinguals wanted their own sponsors. Political groups like the

large cities wanted to become sites. The new groups were accomodated. Sponsors and sites were added in an opportunistic fashion, measurably increasing the political constituency and strength of the program in Congress. The Follow Through budget began to grow.

This caused problems elsewhere, however. Exact comparison groups were difficult to find. Often controls were established that were very unlike the Follow Through classes. The program administrators were aware of these deviations from the evaluation plan, but they felt that the new models would be so much more effective than what the public schools were doing, and the gains in test scores would be so dramatic, that it would not matter whether the comparison classes were closely matched to the Follow Through classes.

Follow Through grew larger and larger. At its zenith there were more than twenty sponsors operating in over one hundred eighty sites. Hundreds of thousands of children were involved. SRI tried to collect data on most of them but the logistics of data collection and the costs bounded out of control. Furthermore, SRI was unable to develop the new instruments it had promised. Amidst investigation by the General Accounting Office, HEW, and Nader's Raiders, the evaluation became a scandal. Finally in 1972, the Follow Through administrators in the Office of Education resigned and the evaluation was reshaped. SRI had spent $12 million on the evaluation in the program's first four years.

Under the direction of the Office of Education the evaluation was pared down to seventeen sponsors working with eighty sites. The analytic sample contained only twenty-thousand children. More importantly, the number of instruments to collect data was narrowed to only four standardized measures. SRI continued to collect the data but the data analysis was contracted to

Abt Associates. In all, the broad scope of the evaluation was drastically narrowed in what Haney (1977) called a "funneling" effect. The early childhood models would now be compared on only a few standardized tests to determine which was "best."

Throughout the course of the evaluation, the sponsors, parents of the children, and site personnel were by no means silent in their objection to events. Most continued to complain, often bitterly, about the evaluation, fearing their models and their children would not be assessed by appropriate criteria. Many never accepted the conversion of the entire Follow Through program to an experiment. They saw Follow Through as a program providing social services to children and their families. Sponsors primarily saw it as a development program.

Faced with the problem of analyzing the test data from nonequivalent (and often mismatched) Follow Through and comparison classes, the Abt Associates analysts resorted to a complex statistical analysis to try to correct the problem as best they could. The technique chosen, analysis of covariance, adjusts the final test scores of children in such a way that their test scores are made more equivalent based upon the achievement test scores of the students, the income level of the parents, and other variables recorded at the beginning of school. Presumably, after the statistical adjustment the scores of the two classes will be more like the scores of two properly matched classes.

Unfortunately, this statistical technique has proved to be much more unreliable in practice than it was believed to be in the late sixties and early seventies (Cronbach, 1977; Campbell and Erlebacher, 1975). The actual test score adjustments are such that the error in the procedure is

quite large. The technique has now become controversial among statisticians. The entire AAI evaluation of Follow Through is based upon i

Another controversial aspect of the evaluation was whether to use individual student scores or class averages in the data analysis. The AAI analysis uses only individual student scores. It has been demonstrated with the Follow Through data that one can obtain dramatically different results using class rather than student scores. Many leading authorities say AAI should have used the class scores instead. This is known as the "units of analysis" problem.

In spite of these and other difficulties, Abt Associates published its results in April, 1977. Based on their controversial techniques, they drew two types of conclusions. One conclusion was that the differences in results from site to site were very great. In other words, even within the same model, e.g. Direct Instruction, many of the sites did better in test results than the comparison classes but in at least two or three Direct Instruction sites the results were much worse than in the comparison classes. This great intersite variation held for all models. The results varied dramatically from site to site for every one.

In fact, the intersite variation among models was so great that the AAI analysts refused to say that any particular model was best. Put in another manner, differences between the sites within a given model were nearly as great as the differences among the models. This was an embarrassment for a study which had been predicated on the idea of identifying the "best" model. In fact, the Office of Education insisted that AAI continue with the comparisons of models in spite of AAI's strong reservation about doing so. In one critique of the evaluation, the finding of great intersite

variation within models seemed to remain valid, despite the many flaws of the evaluation (House, Glass, McLean, and Walker,1978)

The second set of AAI findings revolved around the classification of the early childhood models on the basis of the model's goals. AAI classified models into three types: basic skills, cognitive/conceptual, and affective/cognitive. AAI also classified the outcome measures into basic skills, cognitive, and affective. This dual classification seemed to be extremely arbitrary and perhaps mistaken.

The AAI analysts then matched the appropriate type of model with the corresponding outcome measure. In other words, one would expect the so-called "basic skills" models to do better on the basic skills measures and so on. This gave a semblance of fairness to the evaluation and disguised the fact that the evaluation primarily consisted of standardized achievement tests, the traditional measures on which one might expect "basic skills" models, which emphasized rote learning skills found on such tests, to do better.

The AAI analysts found that the "basic skills" models did better on both basic skills outcomes and on the affective measures. The "cognitive" and "affective" models did better on none of the measures. All this, of course, was within the context of powerful intersite variation, which is to say that any given site from a particular "affective" model might do extremely well on all the measures. For example, although AAI ranked the Bank Street model in the middle in terms of effects, one of its sites was among the best.

In the milieu of the times, the AAI finding that "basic skills" models are better was seized upon by the mass media and the finding of intersite variation was virtually ignored. Articles were carried in the New York Times, Wall Street Journal, Washington Post, Newsweek, and most of the major newspapers in the country. The newspapers were not careful in their coverage of the findings, simplifying the results considerably. Even the AAI analysts were moved to protest the distorted coverage in the Boston newspapers.

Perhaps the most widely circulated report was that of the conservative syndicated columnist, James J. Kilpatrick. In a column that can only be labeled a parody of the AAI report, Kilpatrick wondered why it had taken the educators so much time and money to discover the obvious about schooling. His view of the "basic skills" models was no closer to reality than his description of the "affective" models. His column was widely circulated across the country under various headlines supplied by local newspapers including "Basics Beats Funsies in School," "A Nation of Illiterates," and "Basic Education Offers Alternatives to Numbskulls."

In response to protests about the evaluation by sponsors and others, the Ford Foundation funded a third party critique of the evaluation by a panel of evaluation experts. The panel found the intersite variation finding substantially correct. However, the findings comparing the models were invalid. The critique asserted that the evaluation contained a series of errors and inadequacies.

For example, the narrowness of the scope of measurement and its bias towards certain models precluded statements about which models were

best. Some of the instruments had questionable reliability, and the classification of both the models and measures was misleading. Furthermore, the evaluation contained two substantial statistical flaws. When these flaws were corrected, no models or model types proved to be better, even given the traditional measures used as outcome data. Nonetheless, the finding of intersite variation held up under the reanalysis of the critique (For the full version of the critique, see House, Glass, McLean, and Walker, 1978).

A number of major issues were raised by the Follow Through evaluation: The Big Bang Theory. The idea that one can invent a model program that will "solve" the problems of disadvantaged children across the country was a strong element in the Follow Through program. That belief now seems to be dissipating slowly but steadily. The originators of the program, and possibly the sponsors, thought they could invent educational treatments that would be far superior to the schooling traditionally offered by the public schools. Such dramatic gains were not forthcoming.

It may be that any gains were severely masked by the narrowness of the traditional tests used to measure outcomes of the early childhood models. As a whole, though, the data did not show dramatic results. In the AAI evaluation the Follow Through models as a whole did no better than did the public school classes to which they were compared. It should be noted that the comparison classes themselves were often classes enriched by Title I and other compensatory programs.

Faith in Testing. The faith in testing was strong. Government planners never wavered from the view that gains in scores on standardized achievement tests were the improvements they wanted, whatever else they got. To them the gains in test scores were surrogate measures for improved chances later in life. They insisted that test scores be the focus of any evaluation.

When the sponsors protested that traditional tests were too narrow to measure their program outcomes, Stanford Research Institute expressed confidence that they could develop new measures to cover these new outcomes. They tried but failed miserably. At the end SRI questioned their faith in their own ability to develop such measures and cautioned other test developers.

Most of the sponsors felt that the tests were invalid for their models and protested vociferously their exclusive use. Yet they persisted in the program, hoping against hope, that in spite of poor tests, their own models would show up well on them. Faith in their own models led them to believe they would do well on the tests.

Teaching to the Test. A familiar issue was raised by the evaluation—teaching to the test. It is clear that teaching the exact items on a test is an illegitimate activity--unless the items taught comprise the universe of things to be learned. Most tests, such as standardized achievement tests, only sample the domain of learning that is being assessed. Teaching the items invalidates the inference that the student knows the domain of knowledge the test is sampling.

Other than teaching the items, there are a number of things one might do to prepare for the test, however. It is about the legitimacy of those

activities that people disagree. At the beginning of Follow Through, it was certain that an achievement test would be an important component of the evaluation. Reportedly, one sponsor said, "We don't care what the test is. Just tell us what it is and we'll teach to it" (Egbert, 1971). Other sponsors thought this was an infringement.

No matter what various sponsors did once they knew the test was the Metropolitan Achievement Test (MAT), a test readily available for inspection, it was true that sponsors did best when their curriculum materials come closest to the specific subtests. For example, the strongest performance on any subtest was turned in by Direct Instruction children on the "language" subtest of the MAT. The language subtest consists of a section in which the students discriminate between incomplete sentences and "telling" and "asking" sentences. The other section calls for identifying errors in capitalization, punctuation, and usage.

It was on this subtest that Direct Instruction turned in by far the strongest performance of any sponsor on any test. In fact, this high score accounts for much of Direct Instruction's effectiveness on basic skills since the language subtest was included in basic skills measures. A comparison of the subtest of the MAT with the third grade lessons in Distar, the commercial version of Direct Instruction, shows a close similarity between the two. The format and instruction, though not specific items, were the same. Distar children are repeatedly drilled on content similar to the test. Is this teaching to the test? Different people give different answers. Other sponsors may have geared some of their materials towards the test too. We point out only one example.

Measurement Problems. The evaluators were unable to come up with anything resembling a satisfactory instrument to measure the less tangible outcomes of the models. The two affective instruments had serious deficiencies. Other studies were attempted, such as assessing the impact of the models on the communities in which they were implemented. These were often dismissed for lack of sufficient reliability. Information was ultimately limited to students filling out pencil and paper tests. The interviews with parents and questionnaires to teachers were not fully treated in the final evaluation report. The focus of the evaluation was exceptionally narrow.

Intersite Variation. Even within the traditional tests employed there was enormous variation in outcomes between sites within the same models. This was the one constant finding of the study. Local circumstances--parents, teachers, peers, school environment, home environment--seemed to have a great effect on the classroom outcomes. Even where the early childhood models had their greatest effect on test achievement, their influence was very modest, less than ten percent of the overall variation in test scores. Local conditions had a far stronger impact. This finding raises questions about both the efficacy and type of federal intervention in local districts. According to this study, government programs are very limited in their power to affect traditional outcomes like test scores.

Implementation. Uncertainty has always existed as to how fully the early childhood models were implemented in the school sites. If any situation is advantageous for model implementation, certainly Follow Through should have been. The sites volunteered to work with the sponsors. Both sites and sponsors were paid a substantial sum of money to implement the models. Sponsors worked with a limited number of sites, usually less than ten, over

a long period of time, most for many years. Several cohorts of children went through the models.

Nonetheless, there is evidence that the implementation was not perfect. Early observation studies found differences in implementation across classrooms within models (Stallings, 1975). Furthermore, some sponsors no doubt implemented their models more successfully than did other sponsors.

Statistical Problems. The evaluation raised highly technical but important issues about the statistical analysis employed. The study demonstrates the limits of analysis of covariance techniques, used in this case beyond their capabilities. The implications are that certain types of studies, such as those with non-equivalent control groups, should not be conducted since the statistical corrections cannot be made reliably. A great number of studies are of this type.

Second, the evaluation demonstrates without doubt that the unit of analysis employed--the individual student, the class, or the school district--has a dramatic effect on the results. The selection of the unit becomes a major problem in the design of most studies. In most cases the classroom is probably the appropriate unit of analysis.

Large Scale Experiments. This evaluation throws into question the utility of all large scale experiments. The costs were exorbitant. The evaluation alone cost nearly $50 million. The information gained was not worth the cost. The idea that one can definitively determine the answer to major questions, such as which is the best model of early childhood education is dubious. A series of small studies that contribute to developing knowledge over a period of time is more informative than a massive one. The expectation

that such a large experiment will resolve major problems is unrealistic.

Fairness. The sponsors were promised early in the evaluation that the less tangible outcomes of their models would be measured. The evaluation was unable to deliver on this promise. By the time the true nature of the evaluation became clear, the sponsors were heavily invested and entrenched in sites. Thus the evaluation agreement was unfair to the sponsors. Much of the sponsor's sense of moral outrage can be attributed to a feeling of being treated unfairly. Evaluators should not make promises they cannot keep. At the very least they should renegotiate the understanding between themselves and those being evaluated.

Press Coverage. The interpretation of the Follow Through evaluation was significantly affected, even distorted, by the mass media. In fact, most people's perceptions, even professionals, were shaped by the press coverage rather than the actual study. This is a serious problem. What the press seems to do is to feed stereotypes that they think the public already has. This is the line of easiest and most succinct communication for them. Unfortunately, it also distorts the messages conveyed.

In this case, the press seized upon the "basic skills" label and read their own meaning into it. Since the public was concerned about "back to basics," the Follow Through evaluation was fodder for that particular movement. In the AAI study reading comprehension was not included as a "basic skills" measure but as a "cognitive" measure. Few parents would want "basic skills" that exclude reading. Yet the press seized only the label itself, a label supplied by the AAI analysts.

The damage done by such a misinterpretation is almost impossible to reverse. The mass media are not interested in corrections of yesterday's headlines. The fault is not entirely the media's since only a few people could understand the AAI report as it was written--two thousand pages of statistics and tables. The media may seize on the simple, the familiar, and ignore the complex in self-defense. But the potential for misinformation in the interaction between expertise and the mass media is formidable. In the case of Follow Through it materialized.

Government Policy. Finally, one must question the government policy that shaped the Follow Through evaluation. Policy based on the "big bang" theory assumed there could be an invention or discovery of a model program that would solve the problem of educating disadvantaged children in the early years of school. Further, it assumed that models could be successfully implemented and uniformly effective under any number of differing local conditions. These assumptions seem considerably more dubious today than when Follow Through began.

That is not to say that the federal government should not develop new programs. But one might expect the new programs to have differential effects in different settings. The same program may be desirable in one place and not in another, desirable even for one group in the same place and not for another. It does mean that the government should not propagate one "best" model to the exclusion of all others, since the effects may differ, if in fact they are even ascertainable.

A second reform is necessary in the federal government's evaluation policy. Federal evaluation policy for the last decade has been built on certain assumptions manifested in the Follow Through evaluation. It has

assumed that there is agreement on program goals and in the outcome measures, almost always test scores, on which the programs are to be assessed. It also assumes simple cause and effect relationships.

There are places when such assumption are valid, where such approaches to evaluation will work, but the United States as a whole is not one of them (House, 1978). In a society as pluralistic as the U.S., people often disagree on goals for schooling. They certainly disagree on outcome measures by which to assess programs. And the cause and effect relationships in the social sciences are exceedingly complex. Perhaps the final judgment is that the Follow Through evaluation was entirely too simple for the context in which it was employed.

The material in this publication was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgement in professional and technical matters. Prior to publication, the manuscript was submitted to the Area Committee for Early Childhood Education at the University of Illinois for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either the Area Committee or the National Institute of Education.

REFERENCES

Abt Associates, Inc., Education as Experimentation: A Planned Variation Model, Volumes IV A-D, Sponsored by OE/DHEW, Cambridge, Mass.: Abt Associates, Inc., 1977.

Campbell, Donald and A. Erlebacher, "How Regression Artifacts in Quasi-Experimental Evaluations in Compensatory Education Tend to Underestimate Effects," in C. A. Bennett and A. A. Lumsdaine (Eds.), Evaluation and Experiment, New York: Academic Press, 1975, pp 195-285.

Cronbach, Lee J. with D. R. Rogosa, R. E. Floden, and G. G. Price, "Analysis of Covariance in Nonrandomized Experiments: Parameters Affecting Bias," Occasional paper of Stanford Evaluation Consortium, Stanford University, 1977.

Egbert, Robert, Interview conducted at the Center for Instructional Research and Curriculum Evaluation, April 28, 1977.

Elmore, Richard F., Follow Through: Decisionmaking in a Large Scale Social Experiment, Unpublished doctoral dissertation, Harvard University, 1976.

Haney, Walt, A Technical History of the National Follow Through Evaluation, Sponsored by OE/DHEW, Cambridge, Mass.: Huron Institute, August, 1977.

House, Ernest R., Gene V. Glass, Leslie D. McLean, and Decker F. Walker, "No Simple Answer: Critique of the 'Follow Through' Evaluation," Harvard Educational Review, May, 1978.

House, Ernest R., "The Objectivity, Fairness, and Justice of Federal Evaluation Policy as Reflected in the Follow Through Evaluation," Mimeo, Center for Instructional Research and Curriculum Evaluation, 1978.

McLaughlin, Milbray W., Evaluation and Reform: The Elementary and Secondary Education Act of 1965/Title I, Cambridge, Mass.: Ballinger Publishing Company, 1975.

Stallings, Jane, Implementation and Child Effects of Teaching Practices in Follow Through Classrooms, SRCD Monograph, 40, 7-8, 1975.

Rivlin, Alice, Systemmatic Thinking for Social Action, Washington, DC: The Brookings Institution, 1971.